DOCUMENT RESUME

ED 422 384 TM 028 946

AUTHOR Althouse, Linda Akel; Stritter, Frank T.; Strong, Douglas E.; Mattern, William B.

TITLE Course Evaluation by Students: The Relationship of Instructional Characteristics to Overall Course Quality.

PUB DATE 1998-04-00

NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS *College Students; *Course Evaluation; Higher Education; *Instructional Effectiveness; Medical Education; *Medical Students; Prediction; Science Instruction; Sciences; *Student Evaluation of Teacher Performance; Tables (Data)

ABSTRACT

Student ratings of teaching effectiveness and course quality are widely used to evaluate most university courses. This study investigated the relationship of medical students' ratings of basic science course characteristics to their overall evaluation of these courses. Whether the relationship was similar across basic science courses, and whether course year has an impact on the instructional characteristics students used in making their judgments of overall course quality were also studied. Approximately 160 first-year and 160 second-year medical students were asked to evaluate basic science courses (14 first-year and 16 second-year), for a total of 3,117 individual evaluations made with an instrument developed for the curriculum evaluation process at the medical school. Sixteen instructional items were examined as predictor variables. While 10 of the 16 predictors emerged as significant characteristics using a hierarchical linear modeling calculation, the most dominant predictors clearly were students' ratings on the administrative aspects of a course and the quality of the lectures. Following closely behind these items was the extent to which students felt the course promoted active learning. Other variables also asserted significant influence, although different predictors emerged for each course. These finding suggest that no one student rating item will be useful for all purposes, although when they are combined, a sense of overall course quality can be obtained. (Contains 5 tables and 15 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 422 384

Course Evaluation by Students: The Relationship of Instructional Characteristics to Overall Course Quality

Linda Akel Althouse
Columbia Assessment Services, Inc.

Frank T. Stritter, Douglas E. Strong, William B. Mattern
University of North Carolina at Chapel Hill

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Linda Althouse
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Paper presented at the Annual Meeting

of the

American Educational Research Association

San Diego, California
April 13-17, 1998

TM028946

Course Evaluation by Students: The Relationship of Instructional Characteristics to Overall Course Quality

Linda Akel Althouse
Columbia Assessment Services, Inc.

Frank T. Stritter, Douglas E. Strong, William D. Mattern
University of North Carolina School of Medicine

Introduction

Student ratings of teaching effectiveness and course quality are widely used to evaluate most university courses. Administrators use results of such evaluations in making decisions regarding instructors' promotion, tenure, and salary. Instructors use them in improving their course design and instruction. Students use them to reflect on their overall satisfaction with the school and the educational experience. Researchers use them to identify effective ways of teaching and learning. Because of their widespread use and their importance in decision making, student evaluations of teaching effectiveness have been referred to as "the most thoroughly studied of all forms of personnel evaluation, and one of the best in terms of being supported by empirical research" (Marsh, 1984; Marsh, 1987, p. 369). Extensive research has shown that student evaluations are widely used, reliable, stable, and valid (Abrami, d'Apollonia, & Cohen, 1990; Feldman, 1988; Marsh, 1987; Murray et al., 1990; Neumann & Neumann, 1983). However, most of the literature regarding student evaluation of effective instruction was found to focus on teacher behaviors rather than on course characteristics or quality (Abrami, d'Apollonia, & Cohen, 1997).

In addition, while a number of studies consider the determinants of students' instructional evaluation across widely differing academic areas (Feldman, 1978; Marsh, 1984), no study was

found which reviewed instructional evaluation across specific courses within a particular academic area, such as medicine. In fact, Abrahams and Friedman (1996) noted that despite the prevalence of the use of course evaluations by medical schools, the medical education literature on course and curriculum evaluation is sparse. However, in order to better evaluate curricula and provide faculty development, medical educators need to know the instructional characteristics that contribute the most to students' perceptions of overall course quality. Therefore, the objective of this study was to investigate the relationship of medical students' ratings of basic science course characteristics to their overall evaluation of those courses. The specific questions guiding this study were:

1. What is the relationship of medical students' ratings of basic science course characteristics to their overall evaluation of the quality of those courses? That is, which instructional characteristics predict medical students' overall evaluation of the quality of their basic science courses?
2. Is this relationship similar across basic science courses?
3. Does course year have an impact on the instructional characteristics that students use when making their judgement of overall course quality?

Method

Subjects

This study was conducted at a large state supported medical school on data collected at the end of 1995-1996 academic year. Approximately 160 first and 160 second year medical students were asked to participate in a course evaluation at the end of each basic science course.

As part of the standard course evaluation practice, 30 courses are rated by a questionnaire composed of a common set of seventeen items. The number of students completing the questionnaire for each course ranged from 56-152. A total of 3117 individual evaluations containing the common 17 items were completed.

Instrument

The questionnaire was developed by a faculty committee charged to evaluate the curriculum and based on issues it felt were important in making decisions about the curriculum. The questionnaire consisted of two principal sections. The first section included 16 specific items on instructional design which were rated on a 3, 4 or 5 point Likert scale. These items were used as the independent variables in our study. The second section consisted of a single question that asked the students to rate the overall quality of the course on a 5-point scale. This item was used as the dependent variable. The Cronbach alpha reliability estimate for the instrument was .82. A listing of the items and rating choices on the questionnaire is provided in Table 1

Procedure

The data for the study were collected as part of the standard curriculum evaluation process where students are asked to complete a course evaluation at the end of each of their basic science courses. Therefore, no special data collection effort was needed. Once collected, the data were analyzed in two stages. In the first analysis, each of the 30 courses was considered individually to see which students' instructional design ratings related to students' perception of overall course quality for each specific course. For each course, the set of 16 instructional design ratings was entered into a multiple regression equation. The regression allowed us to determine

the amount of variance that these variables explained in students' rating of overall course quality and which of the variables were significant predictors of the overall ratings. Specific attention was given to the courses that received the lowest and highest overall rating to determine if the same variables emerged as being significant in predicting overall course satisfaction.

In the second analysis, the data from all the basic science courses were combined. This combined data created a hierarchical, multilevel model with students (level 1) nested within courses (level 2). As with regression, all models were fit using listwise deletion for missing data. Since some items did not apply to some of the courses, these items were left blank when the students completed the evaluation. Such missing data values resulted in the loss of three courses (courses 110, 217, and 221) from the combined data set. Therefore, the combined data included only 27 of the 30 courses. Thirteen of these courses were first year course, while the remaining 14 were second year courses.

Since the data were collected at the end of each course within the same academic year, most students completed an evaluation form on more than one occasion. Therefore, when the data were combined the measurements could not be considered independent. However, independence of observations is a necessary assumption for regression. Violation of this assumption could lead to misestimated standard errors. In fact, the combining of the data yielded both course specific (e.g. course 1 is easy, but course 2 is difficult) and student-specific (e.g. student 1 is an easy rater in all courses) error. Therefore, to analyze the effects of the student rated variables on ratings of overall course quality on the combined data, hierarchical linear modeling (HLM) techniques, which is designed to control for these types of errors, rather than multiple regression, were used on the combined dataset. However, since the evaluations were

anonymous, student identification was not available making it impossible to correct for student-specific error.

Three series of multilevel (two levels) models were estimated by use of the HLM (Bryk et al., 1986) program. For each of these series, level one of the multilevel models was the within-unit, student-level model for overall course rating. Level two was the between-unit, course level model.

The first series of models (Model 1) considered only the overall quality ratings for each course and no information on the sixteen instructional design items. This series was equivalent to conducting a random-effect ANOVA. This multilevel model is commonly referred to as the unconditional model and is typically used as a preliminary step in hierarchical data analysis as it provides information on the amount of between-course and within-course variation in overall course ratings. In addition, it provides information about the reliability of each of the course's sample mean as an estimate of its true population mean.

The second series of models (Model 2) added the sixteen student rated items, centered by the course mean, to level one (within-unit level) of the baseline model. Level two of this model was constructed so that the slopes for each of the sixteen items were fixed which meant that the effect of each of the items was assumed to be equal for all courses. This model allowed us to determine the impact of each of the items on the overall quality course ratings by regressing student rating of overall course quality on the sixteen predictor items while controlling for course effect. In addition, it allowed us to determine the amount of variation in overall course ratings which can be explained by this set of items.

The third series of models (Model 3) added the course year, first or second, to level two (between-unit level) of the model used in the second series. This addition allowed us to determine

the effect that the course year had on the ratings given by students. In addition, it allowed us to determine the difference in impact that the sixteen instructional items had on overall course quality ratings for courses taught in the first year versus the second year.

Results

The overall mean ratings and number of completed evaluations for each of the 14 first year courses and 16 second year courses are presented Table 2. These means are listed in descending order based on their overall quality rating. The average overall mean rating for the first and second year courses were 3.556 (sd=.56) and 3.561 (sd=.46), respectively. Since the survey was anonymous, we were unable to analyze the differences between respondents and non-respondents within each course. However, we found no significant relationship (r=.26) between the number of completed evaluations and the overall course quality rating.

Individual Multiple Regression for the 30 Courses

The results of the multiple regression for each of the 30 courses are provided in Table 3. The amount of variance explained by the 16 predictors in the students' overall rating of course quality ranged from 43% to 86% for each of the 30 courses. There did not appear to be a difference in the amount of variance explained between the highest and lowest ranking courses, 71% versus 68%, respectively. The significant ($p<.05$) predictors for the lowest ranking class included quality of the course syllabus, quality of the large group lectures, and the opportunity for interaction with faculty. The significant predictors for the highest-ranking class included administrative aspects, integration of subject matter, quality of large group lectures, and the extent to which small groups illustrated clinical applications.

HLM Analysis of the Combined Data (27 courses)

Prior to beginning the HLM analysis of the full combined dataset, the correlation between each of the indicator variables and the student's rating of overall quality were examined. The linear correlations between these variables and overall course quality rating, as well as descriptive data for each item, are presented in Table 4. The magnitudes of the correlations ranged from .003 to .659. Therefore, no single indicator variable explained more than 43% of the variance of another. Also, most of the high correlations involved the dependent variable indicating that a strong relationship existed between many of the 16 predictors and overall course rating as seen in the HLM results.

The HLM results for the three models using the combined dataset are presented in Table 5. The preliminary HLM analysis of Model 1 found that prior to controlling for the students' ratings of the instructional characteristics, 25% of the total variance in overall course ratings is between courses while 75% of the variance is within courses or at the student level. Therefore, most of the variation in overall course ratings is at the student level. However, a substantial and significant proportion of the variance is between courses. In fact, a significant difference (χ^2= 804.417, df=26) was found in the mean overall ratings of the courses indicating a group effect and confirming the need to perform HLM rather than using a regression analysis. The reliability of the estimate of the mean overall course's sample rating as an estimate of its true population mean rating was .96 which indicated that the sample course quality ratings were reliable indicators of the true overall rating.

When the 16 student rated items were added to the level 1 equation to create the second multilevel model (Model 2), the following student rated variables were found to significantly ($p <$

.01) relate to overall course quality rating: administrative aspects, subject matter integration, quality of course syllabus, fair exams, level of faculty expectation, quality of the lectures, quality of the small groups, coordination with other courses, promotion of active learning, and the number of classes attended. Items not significantly related included: amount of time spent in lectures, amount of time spent in small groups, and the extent to which small groups provided opportunities to collaborate with other students, to interact with faculty, illustrate clinical applications of the basic sciences, and the extent to which they provided hands-on experience that complements the lecture. Highly significant differences (χ^2= 2108.716, df=26) were also found among the 27 course means. By comparing this model with the preliminary model (Model 1) , we found that the 16 student rated items accounted for 62% {(.784-.299)/.784 = 61.9} of the student-level variance in overall course ratings. Of the items that were significantly related to overall course rating, all were positively related except for the level of faculty expectations as evidenced by course exams. If the faculty were viewed as being overally ambitious in their course exams, then students tended to rate them poorly.

In the final HLM model (Model 3), course year was added to the level two between-unit model in order to determine the effect course year had on the ratings given by students. As can be seen in Table 5, course year was not significantly related to the overall course mean rating (p = .970). In fact, course year was only significantly related to two of the student rated items: amount of time spent in small groups and whether small groups valued student collaboration. When Model 3 is compared to Model 2, we found that the addition of the course year resulted in less than one percent of the variation of the student-level variance in overall course ratings. Given that the addition of course year did not add to the explanation of variance between or within groups, we chose to focus our discussion on Model 2 when discussing the impact of each of the

16 predictor variables on overall course rating. However, we did make an exception with the two items where course year did make a difference in Model 3.

Discussion and Educational Implications

While 10 of the 16 predictors emerged as significant characteristics using the HLM calculation, students' ratings on the administrative aspects of a course and the quality of the lectures were clearly the most dominant predictors. The influence of these two characteristics was not a surprise as the literature on teacher evaluation frequently cites these two characteristics as being influential in assessing the effectiveness of a teacher and/or course (Costin, 1971; Feldman, 1997; McKeachie, 1979). In addition, these results agree with the intuitive perceptions of curriculum administrators at this institution, as the results may be reflective of the structure of the present curriculum that is dominated largely by lectures. Also, administrative aspects become very important to students when so much information is presented to them in such a short class time.

Following closely behind these two items was the extent to which students felt the course promoted active learning. This result seems to imply that medical students enjoy and want to take a more active role in learning material. This awareness supports the research reported on the problem based learning methodology which is being introduced across many medical school curriculums. However, when lectures are given, the quality of the presentation directly influences students' ratings of the overall quality of the course. The combination of these characteristic hints to the need and desire for a curriculum that is organized and has high quality lectures with time allocated for active, hands-on learning.

It is important to note that while the above three characteristics had the most influence on perception of course quality, other variables also demonstrated a significant influence attesting to the need that all of these items should be considered in the evaluation of a course. In fact, Cashin (1997) noted that there are several studies that conclude that student rating forms are multidimensional as they measure several aspects of teaching. That is, no one student rating item will be useful for all purposes, but combined we can obtain a clearer sense of overall course quality. In addition, singular items can be used to better understand what contributes to students' perceptions of the quality of a course.

When individual courses were considered, different predictors emerged for each course. This was anticipated since the courses, particularly within the medical school curriculum, have widely varying characteristics, which could account for some of the differences found between the lowest and highest rated course. The items that were found significant most often across the 30 courses included the extent that the course engaged the students in active learning (16 of 30), quality of the lectures (14 of 30), and administrative aspects of the course (12 of 30). This agreed with the three predictor variables found to be influential when the data were combined across all the courses. Only one item, amount of time spent in lectures, was found be insignificant for all 30 courses. This implied that students did not object to spending time in lectures as long as the quality of the lectures was high. While the quality of small groups was found to be significant in 9 of the 30 courses, the quality of specific group activities was rarely significant (3 or less of 30). This result possibly indicated that it did not matter what the small group activities were as long as the overall quality of the activities was high. One other interpretation is that the items now asked about the small groups, are not the key ones, and that students are determining the overall quality of small group activities by other qualities currently not included on the evaluation instrument.

The year that a course occurred in the curriculum was not significantly related to the overall quality rating of a course. In fact, course year was significantly related to only two of the predictor variables: time spent in small groups and the extent to which small groups encourage collaboration with other students. These variables were stronger predictors for first year students indicating that as students progressed through the curriculum, they placed less emphasis on small group activities when forming their opinions about overall course quality. However, more importantly, course year explained less than one percent of the variation of the student-level variance in overall course ratings, indicating a low effect on the determinants of students' perception of overall course quality. This implies that the same evaluation instrument is relevant for both first year and second year courses.

It is necessary to note that while this study addresses the characteristics that students perceive to influence overall course quality, a highly rated course does not guarantee a higher level of learning. Similarly, we do not know if the same instructional characteristics predict the amount of learning that takes place in a course. We also need to emphasize that students' ratings of their medical courses do not provide a complete assessment of a course or the quality of teaching within the course. For example, the amount of material that is retained by students and peer reviews are just two other viable sources of information. Finally, we must take note that while the sixteen characteristics were designed to be uniform for all 30 courses, there are many differences between each of the courses. For example, it is common within medical school to have a course director and then multiple course instructors. However, some courses are taught with only one instructor. In addition, as noted earlier, some courses may stress small group activities while others function entirely as a lecture course. Some courses may be more problem-based while others follow the more traditional lecture modality of teaching. Future research needs

to be conducted to determine the effect of varying course structures, in addition to the sixteen instructional characteristics considered in this study, on overall course quality within a medical school curriculum. Recognizing that the sixteen characteristics do not provide a full measurement of course quality, we note that they do provide some useful information. In addition, as mentioned earlier, formal student ratings have been shown to provide a reasonable, reliable, and valid way of measuring students reactions.

In summary, certain instructional characteristics of a course, as perceived by medical students, do relate to their ratings of overall course quality. While the list of course characteristics included in this study may only be a subset of all the variables which can influence students' perceptions of course effectiveness and their learning, the results of this study do add to what is understood about effective courses. Faculty interested in increasing students' opinions of their courses should first look at improving the administrative organization of their courses and the quality of their lectures. However, faculty must guard itself into completely using the ratings as bias free results. For example, some of best, yet toughest, material delivered will sometimes be rated low. Yet, it is important that faculty with low ratings critically look at their "problem" areas and see what improvements, if any, can be reasonably made. This will take time and commitment, but the rewards will be reaped when students regard a course as being of higher quality and, perhaps, subsequently, learn more effectively.

Table 1: List of Items and Rating Scales on the Evaluation Questionnaire

Core Evaluation Questions	
I. Organization	
Q1	Administrative aspects of the course 1 - Poor 2 - Fair 3 - Good 4 - Excellent
Q2	Integration of the subject matter 1 - Poor 2 - Fair 3 - Good 4 - Excellent
Q3	Overall quality of the course syllabus 1 - Poor 2 - Fair 3 - Good 4 - Excellent
II. Testing	
Q4	Extent to which exams were a fair representation of course subject matter 1 - Did not seem to agree 2 - Significant Deviations 3 - Only Slight Deviations 4 -Almost Perfect
Q5	Level of faculty expectations as evidenced by course exams 1 - Not ambitious enough 2 - Generally realistic 3 -Overally ambitious
III. Lectures	
Q6	Amount of time spent in large group lectures 1 - Too Little 2 - Slightly too little 3 - Correct amount of time 4 - Slightly too much 5 - Too much
Q7	Quality of the large group lectures 1 - Poor 2 - Fair 3 - Good 4 - Excellent
IV. Small Group Activities	
Q8	Amount of time in small group activities 1 - Too Little 2 - Slightly too little 3 - Correct amount of time 4 - Slightly too much 5 - Too much
Q9	Quality of the small group activities 1 - Poor 2 - Fair 3 - Good 4 - Excellent
Q10	Extent to which small group activities conveyed the value of collaborating with other students 1 - Not at all 2 - Modestly 3 - Significantly 4 - Very Significantly
Q11	Extent to which small group activities provided an opportunity to interact with faculty members 1 - Not at all 2 - Modestly 3 - Significantly 4 - Very Significantly
Q11	Extent to which small group activities illustrated clinical applications of the basic sciences 1 - Not at all 2 - Modestly 3 - Significantly 4 - Very Significantly
Q13	Extent to which small group activities provided hands-on experience that complemented lecture material 1 - Not at all 2 - Modestly 3 - Significantly 4 - Very Significantly
V. Coordination with Other Courses	
Q14	Coordination of this course with other courses in the curriculum you have encountered 1 - Poor 2 - Fair 3 - Good 4 - Excellent
VI. General Issues	
Q15	Extent to which faculty has set goals for the curriculum to diminish excessive reliance on rote memorization and promote more active learning 1 - Not at all 2 - Modestly 3 - Significantly 4 - Very Significantly
Q16	Proportion of scheduled course activities (lectures, labs) attended 1 -Less than 50% 2 - 50-74% 3 -75-90% 4 -More than 90%, but not all 5 - All of them
VII. Overall Rating	
Q17	The overall quality of the course 1 - Poor 2 - Fair 3 - Good 4 - Very good 5 - Excellent

Table 2: Descriptive Statistics by Course

	n	Overall Course Quality Mean Rating	sd
First Year Courses			
Course 101	133	4.444	.783
Course 102	119	4.277	.780
Course 103	149	3.960	.936
Course 104	150	3.940	.647
Course 105	81	3.815	.838
Course 106	90	3.767	.984
Course 107	129	3.752	.884
Course 108	97	3.691	.939
Course 109	125	3.488	.768
Course 110	55	3.255	1.075
Course 111	78	3.167	1.086
Course 112	89	3.034	1.027
Course 113	123	2.683	1.074
Course 114	87	2.517	.951
Second Year Courses			
Course 201	84	4.321	.731
Course 202	88	4.284	.742
Course 203	89	4.112	.665
Course 204	127	4.016	.882
Course 205	78	3.974	.789
Course 206	102	3.647	.940
Course 207	74	3.595	.859
Course 208	80	3.438	.953
Course 209	135	3.400	.899
Course 210	81	3.395	1.008
Course 211	140	3.329	1.021
Course 212	96	3.260	.886
Course 213	76	3.237	.798
Course 214	58	3.190	.888
Course 215	83	3.096	.958
Course 216	98	2.684	.892

Table 3: Regression Coefficients of the 16 Questionnaire Items by Course

	n	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q12	Q12	Q13	Q14	Q15	Q16	R^2
First Year Courses																		
Course 101	124	**.193**	**.190**	.119	.051	-.007	-.046	**.195**	-.067	.124	-.087	-.098	**.165**	.073	.062	.116	.016	.71
Course 102	110	.089	**.195**	.136	.096	-.119	-.091	**.192**	-.027	-.054	.066	-.098	**.352**	-.154	.143	.097	.084	.74
Course 103	142	.049	.164	.056			.102	.125	**-.146**	**.157**	.100	.004	.147		-.015	**.197**	**.141**	.65
Course 104	139	.053	.124	.121	**.156**	-.057	.039	**.160**	.005	-.042	**.143**	.063	-.085	.136	**.264**	**.194**	.105	.57
Course 105	76	.217	-.112	.117	.015	-.151	-.019	**.283**	-.137	.072	-.173	**.274**	-.115	.071	.154	.025	.131	.66
Course 106	76	**.239**	.133	**.191**	.032	-.005	-.077	**.189**	.048	**.270**	-.155	-.020	.059	.015	.118	.170	.007	.68
Course 107	120	**.171**	**.159**	**.165**	.057	-.057	-.067	**.203**	-.031	.031	-.043	.122	.053	-.073	-.025	**.257**	.081	.66
Course 108	84	.140	**.272**	.072	-.035	**-.186**	-.044	-.035	.062	.135	-.084	.000	-.006	.182	.101	**.263**	-.005	.67
Course 109	120	.109	-.005	.050	**.197**	**-.143**	-.061	**.272**	-.027	-.043	.057	.030	.056	.026	.106	**.207**	.119	.64
Course 110	49	.190	-.008	.106	-.188	-.007	.034	**.52**	-.021	-.014	**-.244**	**.241**	**-.253**	.148	.147	**.482**	-.007	.86
Course 111	72	.070	.098	.023	.134	**-.179**	-.039	.087	.102	**.400**	-.007	-.069	-.013	-.008	.017	**.336**	.000	.83
Course 112	82	.140	.044	**.200**	-.002	-.069	.138	.068	.147	.156	-.029	-.040	.023	**.257**	.061	**.255**	.042	.72
Course 113	97	**.284**	.221	.069	.121	-.058	.136	.016	-.131	.107	.075	-.138	.075	-.103	.078	.156	.121	.61
Course 114	78	.110	.077	**.199**	.088	-.172	-.105	**.270**	-.071	.092	-.200	**.357**	.113	-.189	-.124	.126	.180	.68
Second Year Courses																		
Course 201	75	**.251**	.004	**.253**	.192	-.134	-.101	**.275**	-.070	-.014	.018	.048	.138	-.094	.149	-.036	.009	.70
Course 202	79	**.240**	.036	.154	.030	-.029	-.071	.161	.197	.096	.194	.042	-.054	**.298**	-.142	.149	.067	.60
Course 203	80	-.010	.080	.138	.158				.075	.159	.168	-.161	-.037		-.251	**.333**	**.351**	.43
Course 204	115	**.178**	.129	**.154**	.114	-.042	-.076	.092	.073	**.235**	.024	-.002	-.180	**.229**	.022	**.218**	**.188**	.70
Course 205	65	.151	.007	**.247**	.126	.044	-.034	.101	.021	**.273**	-.050	.198	.058	-.280	.166	.183	.137	.79
Course 206	88	**.261**	.015	-.014	**.152**	-.042	.047	**.215**	**.191**	**.263**	.124	.096	-.059	.052	-.041	**.147**	**.197**	.81
Course 207	67	.133	.172	**.272**	-.023	-.037	-.090	.122	**-.235**	.122	.054	.114	.089	-.095	.055	**.222**	.117	.79
Course 208	65	**.145**	**.332**	.065	.040	-.088	-.056	.134	.086	.123	-.012	.010	-.130	.204	.094	**.281**	**.176**	.87
Course 209	126	.108		**.318**			-.018	.060	**-.123**	**.324**	.055	.045	-.008		.075	**.148**	.012	.70
Course 210	71	.090	.103	.156	**.166**	.081	-.006	.049	**.149**	.082	.120	.157	-.201	.044	**.210**	**.191**	.057	.82
Course 211	122	.180	.134	.129	**.150**	-.035	.017	.143	.093	-.022	.038	-.018	.046	.003	**.152**	**.170**	.114	.62
Course 212	92	**.231**	.008	.171	-.122	**-.175**	-.079	**.192**	.104	.016	.148	-.074	.030	.028	.170	.169	.035	.63
Course 213	62	**.262**	.159	.137	.073	-.120	-.064	.166	-.023	.149	-.054	.174	-.238	.127	.069	.203	.122	.72
Course 214	54	.208	.249	.218	.062	-.184	-.013	.052	-.005	**-.358**	-.095	.041	.287	.015	.047	.216	.013	.73
Course 215	54	**.283**	.141	-.020	.081	-.144	-.105	**.282**	-.090	n/a	n/a	n/a	n/a	n/a	.032	.114	.137	.66
Course 216	84	.083	**.310**	.144	.066	-.067	-.019	**.214**	**-.013**	**.359**	-.025	-.037	-.099	.040	.070	.098	-.005	.71

Note: **Bold** items indicate significance at the $p<.05$ level.
n/a indicates that this question was dropped out of the analysis as only a few students responded to it.
Blank items indicate that this question was not part of the questionnaire for the particular course.

BEST COPY AVAILABLE

Table 4: Correlations and Descriptive Statistics of the 16 Questionnaire Items and Overall Course Rating

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17
Q1																	
Q2	.645																
Q3	.558	.544															
Q4	.470	.444	.453														
Q5	-.158	-.121	-.172	-.366													
Q6	-.143	-.159	-.115	-.064	.024												
Q7	.510	.483	.493	.403	-.199	-.171											
Q8	-.037	-.003	-.006	-.108	.060	-.144	-.016										
Q9	.334	.341	.359	.291	-.105	.028	.316	-.223									
Q10	.239	.235	.232	.206	-.046	-.023	.187	-.082	.425								
Q11	.299	.299	.268	.200	-.040	-.026	.248	-.018	.516	.506							
Q12	.349	.344	.333	.275	-.093	-.035	.315	-.084	.468	.359	.526						
Q13	.347	.350	.348	.287	-.081	-.043	.290	-.101	.508	.470	.555	.594					
Q14	.477	.486	.396	.425	-.147	-.101	.460	-.034	.300	.214	.288	.331	.324				
Q15	.204	.259	.238	.347	-.236	.031	.309	-.051	.285	.275	.258	.296	.289	.298			
Q16	.069	.083	.069	.008	-.002	-.150	.162	.042	.092	.116	.072	.031	.073	.073	.094		
Q17	.659	.627	.603	.552	-.275	-.139	.632	-.038	.462	.334	.380	.422	.426	.541	.458	.205	
n	2696	2657	2657	2589	2554	2621	2607	2598	2530	2524	2535	2530	2525	2586	2605	2619	2621
mean	3.103	2.955	2.798	2.921	2.173	3.283	2.820	3.008	2.927	2.819	3.136	3.082	3.028	2.939	2.475	3.940	3.564
sd	.821	.804	.947	.814	.478	.832	.716	.918	.821	.906	.816	.851	.872	.754	.904	.994	1.029

BEST COPY AVAILABLE

Table 5: HLM Results for Within and Between Groups Effects

	Model 1			Model 2			Model 3		
	Effect	se	p	Effect	se	p	Effect	se	p
Fixed Effects - Within-Course									
Overall Mean Rating	2.447	.101	.000	2.448	.101	.000	2.461	.328	.000
Administrative Aspects				.239	.023	.000	.198	.069	.009
Subject Matter Integration				.165	.021	.000	.119	.064	.076
Quality of Course Syllabus				.154	.018	.000	.128	.055	.030
Fair Exam Representation				.129	.020	.000	.110	.062	.089
Level of Faculty Expectation				-.155	.027	.000	-.131	.083	.126
Time in Lectures				-.014	.158	.398	.049	.049	.322
Quality of Lectures				.227	.022	.000	.206	.068	.006
Time in Small Groups				.007	.015	.660	-.101	.045	.033
Quality of Small Groups				.108	.019	.000	.093	.059	.130
Small Group - Collaboration				.006	.017	.739	-.108	.053	.052
Small Group - Faculty Interaction				.026	.021	.214	-.022	.062	.722
Small Group - Clinical Applications				.030	.021	.157	.121	.062	.061
Small Group - Hands-on Experience				.024	.019	.223	.010	.059	.863
Coordination with Other Courses				.107	.020	.000	.112	.063	.085
Active Learning				.196	.017	.000	.231	.052	.000
Proportion of Classes Attended				.108	.014	.000	.042	.044	.355
Fixed Effects of Course Year on Level-1 Variables									
Overall Mean Rating							-.008	.205	.970
Administrative Aspects							.031	.046	.514
Subject Matter Integration							.027	.044	.545
Quality of Course Syllabus							.017	.036	.646
Fair Exam Representation							.013	.039	.744
Level of Faculty Expectation							-.017	.054	.759
Time in Lectures							-.044	.032	.172
Quality of Lectures							.014	.044	.746
Time in Small Groups							.073	.030	.022
Quality of Small Groups							.010	.039	.798
Small Group - Collaboration							.072	.034	.045
Small Group - Faculty Interaction							.036	.043	.413
Small Group - Clinical Applications							-.067	.043	.134
Small Group - Hands-on Experience							-.016	.040	.695
Coordination with Other Courses							-.003	.040	.942
Active Learning							-.024	.033	.480
Proportion of Classes Attended							.043	.027	.126

	Model 1				Model 2				Model 3			
	Est.	χ^2	df	p	Est.	χ^2	df	p	Est	χ^2	df	p
Between Course Variation	.264	804.417	26	.000	.269	2108.716	26	.000	.280	2121.125	25	.000
Within Course Variation	.784				.299				.298			
Reliability of Estimate of Overall Course Mean Rating	.961				.985				.986			
Percent of Variance Explained by 16 Predictors					61.9				62.2			

References

Abrahams M. B., & Friedman C.P . Preclinical course-evaluation methods at U.S. and Canadian medical schools. *Academic Medicine*. 1996; 71(4):371-374.

Abrami, P. C., d'Apollonia, S., & Rosenfield, S. The dimensionality of student ratings of instruction: What we know and what we do not. In Perry, R. P., and Smart, J. C. (eds.), *Effective Teaching in Higher Education: Research and Practice*. Agathon Press: New York. 1997.

Abrami, P. C., d'Apollonia, S., & Cohen, P. A. Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology*. 1990; 82(2): 219-231.

Bryk, A. S., & Raudenbush, S. W. *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage. 1991.

Bryk, A. S., Raudenbush, S. W., Seltzer, M., & Congdon, R. T., Jr. *An introduction to HLM: Computer program and user's guide*. Chicago: Scientific Software. 1989.

Cashin, W. E. Student ratings of teaching: A summary of research. In Feldman, K. A. and Paulsen, M. B. (eds.), *Teaching and Learning in the Classroom*. Ginn Press: Needham Heights, Maine. 1997.

Costin, F. Greenough, W. T., & Menges, R. J. Student ratings of college teaching: reliability, validity, and usefulness. *Review of Educational Research*, 1971; 41(5):511-535.

Feldman, K. A. Course characteristics and college students' ratings of their teachers and courses: What we know and what we don't. *Research in Higher Education*. 1971; 9:199-242.

Feldman, K. A. Effective college teaching from the students' and faculty view: Matched or mismatched priorities. *Research in Higher Education*, 1988; 28(4):291-344.

Marsh, H. W. Students' evaluations of university teaching: Dimensionality, reliability, validity, potential bias, and utility. *Journal of Educational Psychology*, 1984; 76: 707-754.

Marsh, H. W. Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*. 1987; 11:253-388.

Marsh, H. W. Research on college teaching: The historical background. *Journal of Educational Psychology*. 1990; 82(2):189-200.

McKeachie, W. J. Student ratings of faculty: A reprise. *Academe*. 1979; 62:384-397.

Murray, H. G., Rushton, J. P., & Paunonen, S. V. Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology*. 1990; 82(2):250-251.

Neumann, Y., & Neumann, L. Characteristics of academic areas and students' evaluation of instruction. *Research in Higher Education*. 1983; 19(3):323-334.